The EMBO Journal vol.3 no.1 pp.175 – 178, 1984

A comparison of genomic coding sequences for feather and scale keratins: structural and evolutionary implications

K.Gregg, S.D.Wilton, D.A.D.Parry[1] and G.E.Rogers*

Department of Biochemistry, University of Adelaide, South Australia, Australia and [1]Department of Chemistry, Biochemistry and Biophysics, Massey University, Palmerston North, New Zealand

*To whom reprint requests should be sent

Communicated by M.Osborn

DNA sequences have been obtained for embryonic chick feather and scale keratin genes. Strong homologies exist between the protein coding regions of the two gene types and between the deduced amino acid sequences of the keratin proteins. Scale keratins are larger than feather keratins and the size difference is mainly attributable to four 13-amino acid repeats between residues 77 and 128 which compose a peptide sequence rich in glycine and tyrosine. The strong similarities between the two peptide structures for feather and scale in the homologous regions suggests a similar conformation within the protein filaments. A likely consequence is that the additional repeat region of the scale protein is located externally to the core filament. Tissue-specific features of filament aggregation may be attributable to this one striking sequence difference between the constituent proteins. It is believed that the genes share a common ancestry and that feather-like keratin genes may have evolved from a scale keratin gene by a single deletion event.

Key words: keratin genes/feather/scale/homology

Introduction

Avian leg scales and body feathers clearly demonstrate the structural variation which can exist in epidermal appendages and are examples of the diverse functions performed by the group of proteins which are collectively called keratins. (Feather and scale keratins are unrelated to the α-keratins of intermediate sized filaments of epithelial cells and wool.) Despite radical differences in their macroscopic features, these two tissues have many common characteristics. In early development, scales and feathers arise by very similar processes and this has led to the suggestion that feathers may have evolved from the scales of an ancestral species (Spearman, 1964). An evolutionary relationship between the two structures has been inferred from limited comparisons of partial protein sequences (Walker and Bridgen, 1976).

The major structural proteins of feathers comprise a large family of very similar keratins (Walker and Rogers, 1976) and peptide sequence comparisons have shown homologies to exist between feather keratins and the keratins of scale tissue (Walker and Bridgen, 1976). Whether such homologies arose by common ancestry or by convergent evolution is difficult to determine from the peptide sequences currently available. Indeed, extensive peptide sequences are difficult to obtain from complex mixtures of similar molecules such as the chick feather keratins. A more amenable protein population occurs in the emu feather and the most complete peptide sequence so far determined was from that source (O'Donnell and Inglis, 1974). However, recombinant DNA studies have allowed the amino acid sequences of chick feather and scale keratins to be deduced from the nucleotide sequence of their genes (Molloy *et al.*, 1982; Gregg *et al.*, 1983; S.Wilton, in preparation). This detailed information has allowed a direct comparison of chick feather and scale keratins and a comparison of the genes which encode them.

We report here that considerable homology exists between typical members of the embryonic chick feather and scale keratins and that the close relationship of the proteins is reflected in their gene sequences. We believe that these data show that feather and scale keratins arose from a common ancestor.

Results

A composite diagram, shown in Figure 1, illustrates the comparison between the sequences of feather keratin gene B from λCFK1 and scale keratin gene III from the clone λCSK8. Approximately 75% of the feather keratin coding sequence can be matched to sequences within the scale keratin gene with an identical reading frame. Over these regions of the genes, homology is >75% and the amino acid sequence homology is ~70%. A further 9% of the feather coding sequence can be matched to scale sequence, but over these regions the codon reading frames are out of phase and nucleotide sequence homology is only 64%. The remainder of the feather gene coding sequence (~16%) shows no similarity in sequence to the analogous part of the scale keratin gene.

Throughout the two genes, the matching of sequence assumes the occurrence of a total of five deletions and/or insertions. Four of these events are observed as DNA sequences present in the scale gene and absent from the feather gene. The fifth is a 19-base sequence which occurs in the feather gene but not in the scale gene. The most prominent example of this type of alteration is a region of 156 bases which encodes a characteristic part of the scale keratin. This is a series of four 39-base repeats which specify a peptide rich in glycine and tyrosine. Figure 2 shows schematically how the two protein coding sequences are related; indicating how a feather keratin gene might, by a series of deletions, have been derived from an ancestral scale keratin gene.

Regions of directly comparable DNA sequences were calculated to have diverged by 23% in the positions which would lead to amino acid replacement. Assuming that feather keratins appeared before the first known feathered creatures (Ostrom, 1976) some 170 x 10^6 years ago, then it can be calculated that a 1% change has occurred in each 7.4 x 10^6 years.

Discussion

Feather keratins are smaller than avian scale keratins, the two types having mol. wts. of 10 500 and ~15 000, respectively, and the difference is largely accounted for by the four 13-amino acid repeats characteristic of the scale protein. The

© IRL Press Limited, Oxford, England.

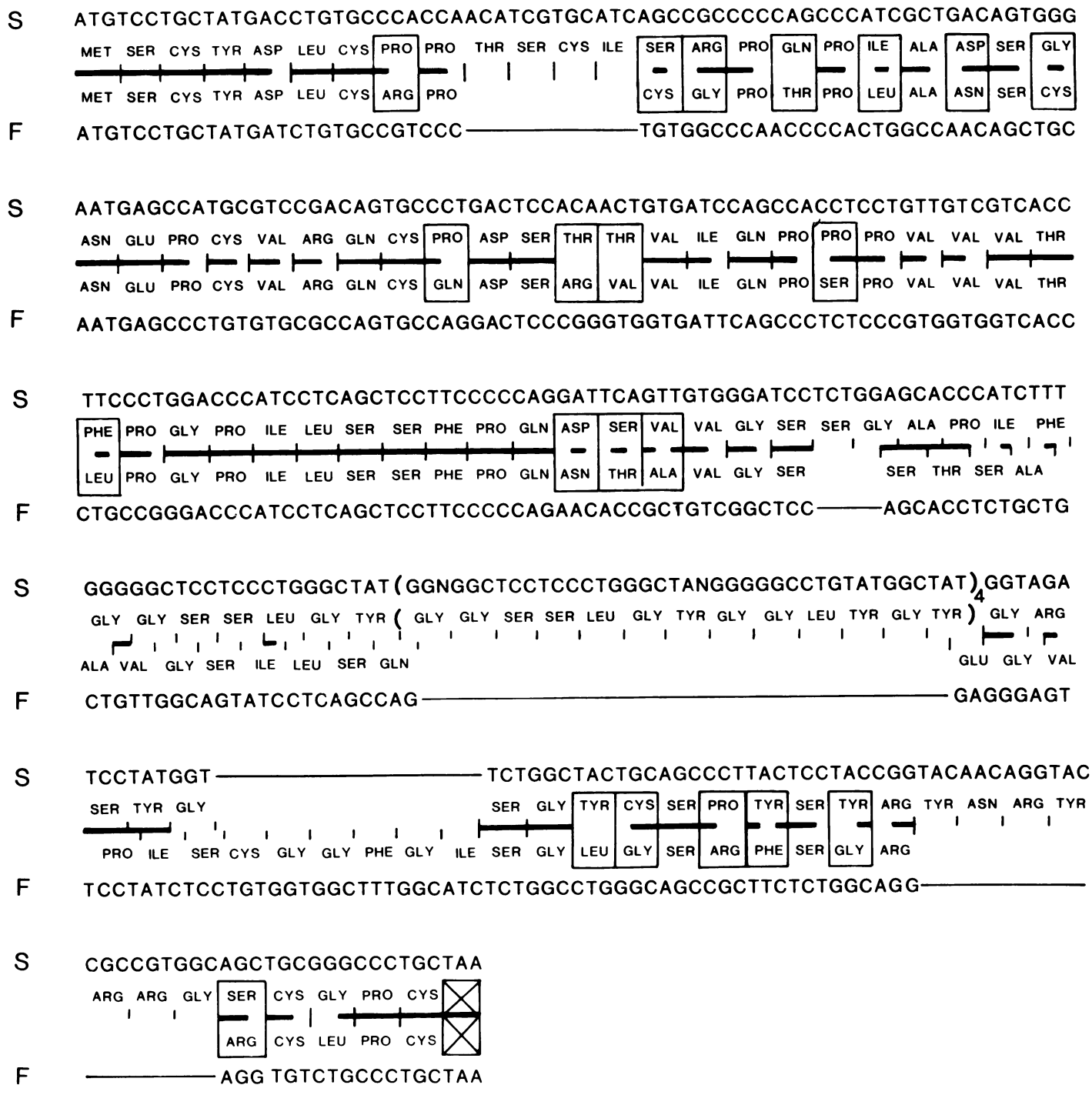

Fig. 1. The complete sequence of the protein coding region of chick feather (F) and scale (S) keratin genes. The dark line between sequences indicates base homology, fine lines within a sequence indicate an apparent deletion. Boxed amino acids are those which differ between the two deduced protein sequences. The region in brackets is the 39-base repeat of the scale keratin gene.

other deletion/insertion events, deduced from matching the nucleotide sequences, do not alter the amino acid content of the proteins to the same extent as this 52-residue alteration. In discussing a comparison of the proteins, it is most convenient to consider separately the similar and dissimilar regions.

The part of the peptide structure which is directly comparable in scale and feather keratin (i.e., scale residues 1–63) is made up of two structurally distinct portions (scale residues 1–27 and 28–63). Structural predictions, using the Chou-Fasman and Robson scheme (Chou and Fasman, 1978, 1979; Garnier *et al.*, 1978) indicate that the region of greatest similarity between the two sequences (i.e., scale residues 28–63) adopts a highly regular conformation which accounts for virtually all of the β-structure present in the segment. On the other hand, the amino- and carboxy-terminal peptides (scale residues 1–27 and 130–155, respectively) represent less structured 'tails' which appear capable of making ionic interactions with one another. These terminal peptides contain essentially all of the SH groups which are ultimately oxidised to give the stabilizing disulphide bonds characteristic of the mature tissue.

The 13-amino acid repeat region is a distinctive feature of

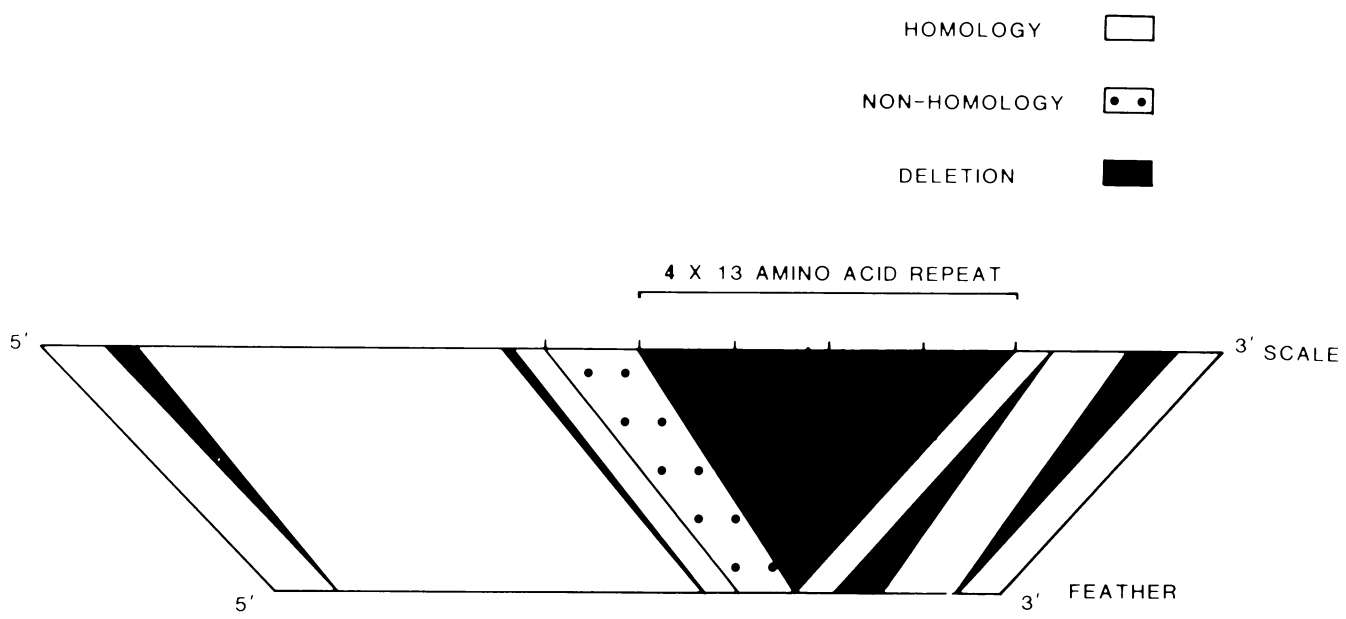

Fig. 2. Schematic representation of the structural similarities between avian scale and feather keratin genes, within the protein coding regions. The regions denoted as homologous have >60% base homology. The non-homologous region (dotted) shows no homology above that expected from random sequence.

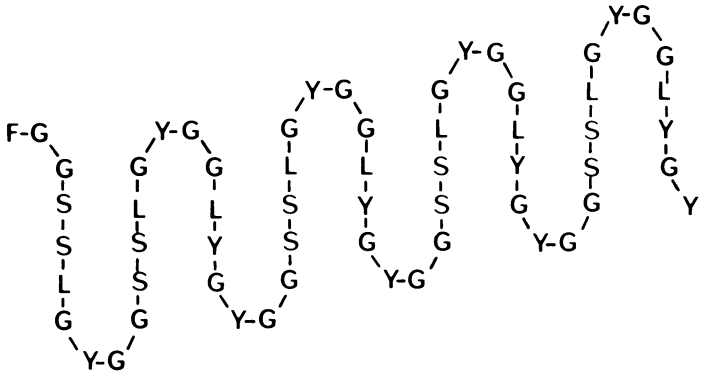

Fig. 3. Diagramatic representation of the potential β-pleated sheet formation within the four x 13-amino acid repeat region of the chick scale keratin.

the scale proteins and is predicted to form an extensive region of alternating β and turn structure, thus leading to a conformation of the type shown in Figure 3. X-Ray studies of powdered material from the scale keratin repeat sequence had previously suggested a β-conformation for this segment (Stewart, 1977) and thus this observation provides strong support for the structure postulated (Figure 3). The scale sequence residues 28 – 63, which is on the N-terminal side of the repeat sequence (Figure 1), is strongly homologous with feather keratin and displays the same eight-residue structural periodicity in the β and turn conformational parameters first described by Fraser and MacRae (1973) for Silver Gull feather keratin. Thus, in scale keratin this segment will also consist of four or five β-strands making up a twisted antiparallel sheet. Two of these sheets will probably aggregate and form the same type of core region typical of the 3 – 4 nm diameter filaments in feather keratin (Fraser *et al.*, 1971). The repeat region of the scale sequence would necessarily be excluded from the compact core region of the filament and would most likely take up a β-conformation on the outside face of the filament. This structure, which would be flanked by rows of tyrosine residues (Figure 3), would be expected to establish strong hydrophobic interfilament interactions. In this respect it is interesting to note that scale filaments are known to form three-dimensional matrices, whereas the feather filaments form predominantly two-dimensional lattices (Filshie and Rogers, 1962; Fraser and MacRae, 1973; Fraser *et al.*, 1971).

The strong homologies which exist between the protein coding regions of feather and scale keratin genes suggest that the two proteins do indeed share a common ancestor. Since the development of scales chronologically preceded the appearance of feathers, it may reasonably be assumed that feather keratins evolved from a scale keratin. At some time after the initial change the feather keratin gene appears to have undergone gene duplication to produce the present family of closely related structures. From the calculated divergence of feather and scale genes, it appears that a 1% divergence occurs each 7.4 x 10^6 years. Assuming that the divergence of feather gene sequences from one another is under similar evolutionary pressure, then the cluster of feather keratin genes in the clone λCFK1 can be calculated to have undergone duplication between 7.4 and 120 x 10^6 years ago (divergences range from 1% to 16%, Gregg *et al.*, 1983). However, such calculations must be viewed with caution because of the occurrence of events such as gene conversion (Slightom *et al.*, 1980) which could reduce the apparent divergence between similar genes.

If it is assumed from the genetic homologies that feather keratins were indeed developed from the proteins of scale tissue, then an important evolutionary question is: how much did the prior alteration of the peptide sequence facilitate the morphological change of the tissue? All feather keratins so far investigated are of a type similar to that found in the chick (O'Donnell and Inglis, 1974) with no significant contribution being made by scale-like proteins. In structural analysis it has been shown that the filaments of feather and scale are of a similar type, although the scale filaments may be more tightly wound (Stewart, 1977). Taken collectively, the molecular

features suggest that the nature of the protein may significantly determine the conformation of filaments and their aggregation. If the nature of the keratin components is important in the assembly of scales or feathers then the major difference in this respect could have occurred in a single event with the deletion of the peptide repeat region. In this way the protein characteristic of scale tissue could have been transformed spontaneously into a typical feather-like keratin. Such a change may be facilitated by the very nature of the DNA sequence since it has been shown that repeat sequences frequently flank DNA elements which are mobile within the genome (Shen *et al.*, 1981). Certainly, repeated sequences provide the opportunity for unequal recombination events leading to the acquisition of additional DNA by one daughter molecule and loss of DNA from the other (Shen *et al.*, 1981). This is one way in which a species ancestral to the first known feathered creature, *Archaeopteryx lithographica* (Ostrom, 1976), could have developed scales with an altered fibre structure; the first step towards the evolution of feathers.

Materials and methods

Keratin gene clones

DNA clones were obtained by screening a Charon 4A chicken genomic library (prepared by J.Dodgson, J.Engle and R.Axel) with cDNA from the mRNA of either 14-day embryonic chick feathers or 17-day embryonic leg scales (Molloy *et al.*, 1982; S.Wilton, in preparation). Further partial sequences were obtained from mRNA derived clones (S.Wilton, in preparation) and served to confirm that the genomic sequences obtained were representative of the appropriate transcripts.

DNA sequencing

DNA sequencing was by a combination of the chemical cleavage method of Maxam and Gilbert (1980) and the dideoxynucleotide chain termination procedure of Sanger *et al.* (1977) after cloning of DNA fragments into M13 phage (Messing *et al.*, 1981).

Sequence comparisons

Feather and scale keratin gene sequences, which were typically representative of their families, were used for the direct comparison of both nucleotide and peptide sequences (Gregg *et al.*, 1983). Calculation of divergence between homologous DNA sequences was made using the amino acid replacement changes as described by Perler *et al.* (1980). Base changes in silent positions were not included in the calculations because of the poor correlations obtained by those authors when examining genes for globin and preproinsulin.

References

Chou,P.Y. and Fasman,G.D. (1978) *Adv. Enzymol.*, **47**, 45-148.

Chou,P.Y. and Fasman,G.D. (1979) *Biophys. J.*, **26**, 367-383.

Filshie,B.K. and Rogers,G.E. (1962) *J. Biophys. Biochem. Cytol.*, **13**, 1-12.

Fraser,R.D.B. and MacRae,T.P. (1973) *Conformation in Fibrous Proteins*, published by Academic Press, NY.

Fraser,R.D.B., MacRae,T.P., Parry,D.A.D. and Suzuki,E. (1971) *Polymer*, **12**, 35.

Garnier,J., Osguthorpe,D.J. and Robson,B. (1978) *J. Mol. Biol.*, **120**, 97-120.

Gregg,K., Wilton,S.D., Rogers,G.E. and Molloy,P.L. (1983) in Nagley,P., Linnane,A.W., Peacock,W.J. and Pateman,J.A. (eds.), *Manipulation and Expression of Genes in Eukaryotes*, Academic Press, Sydney, pp. 65-72.

Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499-560.

Messing,J., Crea,R. and Seeburg,R.H. (1981) *Nucleic Acids Res.*, **9**, 309-321.

Molloy,P.L., Powell,B.C., Gregg,K., Barone,E.D. and Rogers,G.E. (1982) *Nucleic Acids Res.*, **10**, 6007-6021.

O'Donnell,I.J. and Inglis,A.S. (1974) *Aust. J. Biol. Sci.*, **27**, 369-382.

Ostrom,J.H. (1976) *Linnaean Soc. London, Biol. J.*, **8**, 91-182.

Perler,F., Efstratiadis,A., Lomedico,P., Gilbert,W., Kolodner,R. and Dodgson,J. (1980) *Cell*, **20**, 555-565.

Sanger,F., Nicklin,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.

Shen,S., Slightom,J.L. and Smithies,O. (1981) *Cell*, **26**, 191-203.

Slightom,J.L., Blechl,A.E. and Smithies,O. (1980) *Cell*, **21**, 627-638.

Spearman,R.I.C. (1964) in Ebling,F.J. (ed.), *Symposium of the Zoological Society of London No. 12*, Academic Press, London, p. 67.

Stewart,M. (1977) *J. Ultrastruct. Res.*, **60**, 27-33.

Walker,I.D. and Bridgen,J. (1976) *Eur. J. Biochem.*, **67**, 283-293.

Walker,I.D. and Rogers,G.E. (1976) *Eur. J. Biochem.*, **69**, 341-350.

Received on 4 October 1983

Note added in proof

The authors wish to acknowledge financial support from the Wool Research Trust Fund on the recommendation of the Australian Wool Corporation (S.D.Wilton) and from the Australian Research Grants Committee (Keith Gregg and G.E.Rogers).